The EuroBiotech Journal
REVIEW ARTICLE
Medical Biotechnology

Artificial Intelligence and Humanoid Robotics: Bioethical Implications of Replacing Human Agency in Healthcare and beyond

Sara Feizyab[1,*], Gabriele Bonetti[2,3], Maria Chiara Medori[2], Cecilia Micheletti[2], Immacolata De Luca[2], Kevin Donato[1, 4], Jan Miertuš[2], Mehmet Sait Dundar[5], Miroslava Vráblová[6], Gary Henehan[7], Richard Brown[8], Robert Marks[9], Stanislav Miertus[10,11], Lorenzo Lorusso[12], Gianluca Martino Tartaglia[13,14], Munis Dundar[15], Sandro Michelini[16], Stephen Thaddeus Connelly[17], Ariola Bacu[18], Tommaso Beccari[3], Ornela Gordani[19], Xhilda Dhamo[19], Eglantina Kalluci[19], Dominika Vešelényiová[20], Iveta Dirgová Luptáková[21], Jiri Pospíchal[21] , Matteo Bertelli[1,2,4]

Abstract

The integration of advanced artificial intelligence (AI) and humanoid robotics into healthcare represents a critical evolution in biotechnology with profound societal implications. This review explores the bioethical implications of these technologies, and their potential to displace human agency in life-critical decisions. It adopts an interdisciplinary approach encompassing ethics, law, and technology. The review examines how innovations in AI and robotics might shift autonomy from humans to machines, and addresses the accountability challenges inherent in such transitions. We synthesize discussions on the ethical management of AI and robotics, underscoring the importance of maintaining human oversight and integrating ethical standards in technology development to prevent worsening of social inequalities. While AI and robotics present challenges to traditional concepts of autonomy, ethical responsibility, and justice, careful and inclusive policymaking and ethical oversight can harness these technologies to enhance human well-being. This analysis highlights the necessity for continued cross-disciplinary research to navigate the complex ethical landscapes these technologies create, emphasizing that the proactive engagement of diverse stakeholders is essential to guide AI and robotics towards improving human health.

Keywords: Artificial Intelligence; Humanoid Robotics; Bioethics; Social Justice; Autonomy

[1] MAGI EUREGIO, Bolzano, Italy;
[2] MAGI'S LAB, Rovereto (TN), Italia;
[3] Department of Pharmaceutical Sciences, University of Perugia, Perugia, Italy;
[4] MAGISNAT, Atlanta Tech Park, Peachtree Corners (GA), USA;
[5] Medical Imaging Techniques, Halil Bayraktar Vocational Health School,Erciyes University, Kayseri 38039, Turkey
[6] Department of Criminal Law, Academy of the Police Force in Bratislava, Slovakia;
[7] School of Food Science and Environmental Health, Technological University of Dublin, Dublin, Ireland;
[8] Department of Psychology and Neuroscience, Dalhousie University, Halifax, Nova Scotia, Canada;
[9] Department of Biotechnology Engineering, Ben-Gurion University of the Negev, Beer-Sheva, Israel.
[10] Department of Biotechnology, University of SS. Cyril and Methodius, Trnava, Slovakia
[11] International Centre for Applied Research and Sustainable Technology, Bratislava, Slovakia.
[12] UOC Neurology and Stroke Unit, ASST Lecco, Merate, Italy.
[13] Department of Biomedical, Surgical and Dental Sciences, Università degli Studi di Milano, Milan, Italy;
[14] UOC Maxillo-Facial Surgery and Dentistry, Fondazione IRCCS Ca Granda, Ospedale Maggiore Policlinico, Milan, Italy;
[15] Department of Medical Genetics, Erciyes University Medical Faculty, Kayseri, Turkey.
[16] Vascular Diagnostics and Rehabilitation Service, Marino Hospital, ASL Roma 6, Marino, Italy.
[17] San Francisco Veterans Affairs Health Care System, University of California, San Francisco, CA, USA.
[18] Department of Biotechnology, University of Tirana, Tirana, Albania.
[19] Department of Applied Mathematics, Faculty of Natural Sciences, University of Tirana;
[20] Department of Biology, Institute of Biology and Biotechnology, Faculty of Natural Sciences, University of Ss. Cyril and Methodius in Trnava, 917 01 Trnava, Slovakia;
[21] Institute of Computer Technologies and Informatics, Faculty of Natural Sciences, University of Ss. Cyril and Methodius, J. Herdu 2, 917 01 Trnava, Slovakia;

***Corresponding author:**
Sara Feizyab, sara.feizyab@assomagi.org

DOI: 10.2478/ebtj-2025-0019

Introduction

The rapid advancement of AI and humanoid robotics marks a significant shift in 21st-century medical sciences. These technologies extend human capabilities but run the risk of increasingly replacing humans in ethically sensitive domains such as robotic surgery, AI-driven drug discovery, and even genomic diagnostics (1,2). For instance, systems like the da Vinci Surgical Robot or AI platforms for cancer prognosis exemplify how decision-making autonomy is shifting from human clinicians to algorithms, raising urgent questions about accountability, patient safety, and the value of human judgment in life-critical decisions (3). As human–machine interactions become deeply integrated into biomedical contexts, ethical and ontological challenges are becoming increasingly significant. The delegation of cognitive and physical tasks to nonhuman systems raises complex issues related to human autonomy, identity, and control. In domains such as robotic caregiving, autonomous vehicles, and assistive social robotics, the development of ethical accountability mechanisms and the establishment of clear respon-

sibility frameworks are particularly pressing. These developments require a reassessment of traditional concepts of agency and ethical oversight in the light of nonhuman actors performing roles that directly impact human welfare (4). Nevertheless, numerous studies have emphasized the exceptional benefits of AI and robotics, especially in the healthcare sector. For instance, predictive analytics powered by AI has shown strong potential in enabling proactive interventions and resource allocation by identifying disease patterns based on historical and demographic data (5). Furthermore, AI supports personalized medicine by tailoring treatments to individual patient characteristics, thereby enhancing care outcomes and efficiency.

Mainstream discussions of AI and robotics often focus on efficiency, innovation, and scalability, but these considerations can overshadow pressing normative concerns. The deployment of humanoid systems, which integrate technologies such as computer vision, natural language processing, and decision-making algorithms, raises ethical questions regarding justice, agency, and accountability (6). These concerns are particularly salient when such systems operate through architectures outside of complete human supervision, such as deep neural networks (7). Addressing these issues requires more than technical evaluation; it necessitates interdisciplinary analysis that incorporates perspectives from ethics, philosophy of technology, robotics, and law. Ethical assessment must account for both the capabilities of systems such as simultaneous localization and mapping (SLAM, building a map of an unknown environment while estimating real time position), sensor fusion (integrating data from multiple sensors to improve reliability), and reinforcement learning (a machine-learning paradigm in which an agent learns through reward–penalty feedback) as well as the broader socio-political implications, including shifts in labor dynamics, authority, and ethical responsibility (8).

Notably, a tension exists between focusing on future versus present ethical risks of AI. While some futurists warn that superintelligent AI could pose existential risks to humanity, critics of longtermism (the ethical stance that prioritizes actions whose benefits extend to the far future) argue that an exclusive focus on speculative future catastrophes fails to address serious problems happening now and may even serve to justify compromising people's present rights and interests for the sake of a far-off greater good (9). In the context of healthcare, an excessive preoccupation with distant AI can be ethically distracting, diverting attention from pressing justice issues that are real, imminent and pervasive today (10). In this scheme, this paper reviews the call for multigenerational bioethics (an ethical framework that weighs biomedical decisions' consequences for both present and future generations), balancing long-term concerns against urgent present-day ethical challenges (9).

We first examine the technological and historical trajectory that has led to the current role of AI and robotics, highlighting how the displacement of human agency has deep philosophical roots. We then discuss the implications of AI and robotics for labor and the political economy, the challenges of autonomy and ethical responsibility in artificial agents, and the regulatory responses emerging to govern these technologies. Next, we focus on the medical paradigm as a high-stakes context for AI, exploring how autonomous systems are transforming healthcare ethics. Finally, we consider issues of social justice, including how AI and robotics might reinforce or mitigate structural inequalities. Throughout the discussion, we integrate perspectives from key ethical frameworks such as consequentialism, deontology, virtue ethics, and information ethics to analyze how each framework addresses (or fails to address) the ethical challenges posed by intelligent machines. Finally, we conclude by reflecting on the need for ongoing ethical vigilance and inclusive governance to ensure that AI and humanoid robotics remain anchored in human values such as dignity, fairness, and responsibility.

Autonomy, ethical responsibility, and ethical decision-making in artificial agents

As AI systems gain functional autonomy, attributing ethical and legal responsibility for their actions becomes increasingly complex. Intelligent agents are now deployed in sensitive domains such as autonomous vehicles, surgical robots, military drones, and algorithmic decision-making in criminal justice and finance. These systems produce outcomes that directly impact human lives, rights, and well-being. Unlike earlier automated machines (which followed deterministic rules set by humans), many modern AI systems operate with a degree of independence, making real-time decisions in dynamic, uncertain environments and sometimes learning or adapting as they go (11). This raises a foundational ethical question on how to assign responsibility when an autonomous machine causes harm. Potentially liable actors include the manufacturer, the software developer, the end-user, the deploying institution, or, in more abstract terms, the algorithm itself (12,13).

Classical moral theories offer limited guidance here, because they generally assign responsibility only to entities capable of intention and understanding. Deontological ethics (exemplified by Immanuel Kant) restricts moral agency to rational beings who can act from duty and understand the moral law. Virtue ethics emphasizes the cultivation of moral character and practical wisdom (phronesis), traits that are developed through human habit and experience. Current AI systems, no matter how sophisticated, lack consciousness, intentions, and moral awareness; they simulate decision-making without any internal comprehension or volition. From a Kantian perspective, an AI cannot be a moral agent because it cannot reason about right and wrong or be motivated by respect for moral law (14). From a virtue ethics perspective, a robot cannot cultivate virtues or vices since it has no genuine character or life experience. To summarize, outsourcing ethically consequential decisions to non-sentient systems disrupts traditional paradigms of moral agency and accountability (15).

From a consequentialist or utilitarian perspective, the ethical value of AI decisions may be judged solely by their outcomes, regardless of who makes them (e.g. fewer errors, greater overall health benefits). Indeed, some proponents of AI in medicine

suggest that if an algorithm saves more lives through accurate diagnoses or surgeries, a consequentialist calculus would favor its use. However, critics warn that exclusively outcome-focused reasoning can justify unethical means or erode individual rights (10). This is evident in the longtermism debate: maximizing certain future outcomes (like preventing a hypothetical catastrophe) could be used to rationalize harming or neglecting people in the present (9). In the context of AI, a consequentialist approach needs constraints; otherwise, one might ignore issues of consent, autonomy, or justice for the sake of efficiency or aggregate welfare. Thus, most ethicists agree that AI systems, especially in healthcare and public domains, must be evaluated not only by their results but also by the fairness and transparency of their processes.

Technological evolution and the displacement of human agency

Automation has long been a central vector of human progress. Since prehistoric times, humans have engineered tools to enhance their physical strength and capabilities. However, the current trajectory of AI and robotics does not simply accelerate efficiency, it fundamentally alters the structure of agency itself. For the first time in history, cognitive and decision-making processes including medical diagnosis, legal reasoning, emotional care, and strategic planning are increasingly being entrusted to systems devoid of consciousness, intentionality, or ethical sensibility. This marks a paradigm shift in the anthropological foundations of work and social life (16). It is not merely a functional substitution of labor by machines, but a transformation with far-reaching ethical, legal, and existential implications. Robotic systems equipped with autonomous navigation, probabilistic reasoning, and adaptive behavior can now perform tasks that previously required human experience and contextual understanding (17).

When machines take on roles that involve care or judgment, does it diminish the role of empathy, compassion, or human intuition in those domains? In healthcare, for example, if an AI-driven diagnostic system or an autonomous surgical robot makes critical decisions, the traditional patient–caregiver relationship is altered. The human clinician's agency is partly ceded to a device that cannot truly "care" or understand in the human sense. Such displacement challenges long-standing assumptions about human uniqueness and the value of human oversight. Ensuring that these technologies remain tools for humans, rather than replacements of humans, is a core ethical concern. Many argue for maintaining meaningful human-in-the-loop control in life-critical applications so that human judgment can always override or guide automated decisions when necessary (18). In this context, every algorithmic recommendation must remain advisory until the attending clinician validates or rejects it, and any clinician-AI disagreement should pause automation and alert human staff for immediate checks.

Labor displacement and the political economy of automation

Another ethical concern is based on how automation transforms the structure of labor and the distribution of social and economic power. Indeed, one of the most visible impacts of AI and robotics is on the world of work. The sociotechnical consequences of AI-driven automation have been examined across diverse intellectual traditions. Over a century ago, Karl Marx identified mechanization as a source of worker alienation estranging laborers from both their work and its outcomes. In his later notion of the "general intellect," elaborated in Grundrisse, Marx anticipates today's algorithmic infrastructure, an era where centralized, nonhuman systems accumulate knowledge and decision-making power previously vested in human workers (19). Max Weber's analysis of bureaucratic rationalization in The Protestant Ethic introduced the metaphor of the "iron cage", wherein procedural efficiency can override human values (20). Jürgen Habermas, in the Theory of Communicative Action (1981), later underscored the erosion of participatory agency under technocratic systems (21). Together, these thinkers warn that uncritical embrace of technical rationality can encroach on human freedom and meaning.

In the contemporary context, the post-industrial futurist Jeremy Rifkin predicted that intelligent automation would lead to widespread labor displacement and a dual society: a tech-empowered elite versus a marginalized class of unemployed. Such projections are increasingly supported by empirical data (22). For example, the OECD's Employment Outlook (2023) reports that, on average across its member countries, 27% of jobs are at high risk of automation, especially those involving routine tasks in sectors like manufacturing, administrative support, and finance (23). Likewise, the World Economic Forum's Future of Jobs analysis (2020) projected that while 85 million jobs may be displaced by 2025 due to automation and AI, about 97 million new roles could emerge in that same timeframe (24). These shifts demand proactive societal responses: investment in workforce retraining, education in digital skills, and social safety nets for those most affected. Without such measures, there is a risk of exacerbating economic inequality and creating what Yuval Noah Harari has termed the "useless class" of individuals rendered unemployable by technology (25).

From a bioethical perspective, the loss of meaningful work is not only an economic problem but also a challenge to fundamental human needs. Philosophical traditions from Aristotle to Hannah Arendt have emphasized that labor is not merely a means of survival, but a vital component of human identity, ethical development, and participation in communal life. Arendt's analysis in The Human Condition emphasizes that work and action confer purpose and a sense of belonging in society (26). If AI-driven automation erodes opportunities for people to contribute meaningfully through work, it raises the question of how can human dignity and agency be preserved in a highly automated society? This question intersects with bioethics when considering technologies like care robots or AI in healthcare administration, where work is tied deeply to empathy and human connection. Ensuring that automation does not strip away the dignity of work or reduce persons to passive recipients

of algorithmic decisions is a ethical imperative. It calls for an ethical evaluation of innovation not just in terms of productivity, but in terms of its impact on human flourishing (27).

Legal responsibility and accountability

Although AI offers substantial benefits in diagnostics, drug discovery, clinical management, and surgery, evidence suggests significant challenges related to errors and accuracy (28,29). AI applications have demonstrated promise in predictive analytics, individualized treatments, and efficient resource allocation in healthcare (5,30). Yet, despite these advancements, studies have identified concerning limitations, such as inconsistent performance, opacity of algorithms (when an algorithm's internal logic is difficult to interpret), and uncertain error rates (30,31).

Critical evaluations reveal that AI-generated content can be inaccurate, fictional, or supported by non-existent references, highlighting risks associated with blind trust in these technologies (31,32). For instance, Obermeyer et al. showed that a population-health algorithm widely deployed in U.S. hospitals sent disproportionately fewer Black patients to high-risk care programs because it used past health-care spending as a proxy for medical need, thereby embedding racial bias into triage decisions. Jin et al. illustrate that a leading LLM correctly answered 81.6 % of multiple-choice clinical questions but only 27.2 % of image-based cases, and 35.5 % of its correct answers were justified with flawed clinical rationales(33–35). Moreover, the error rate in AI-driven healthcare remains inadequately quantified across specialties (36,37). Accountability for errors and the potential deskilling of medical professionals further complicate ethical and legal considerations (38). Recent studies indicate that combining AI with human expertise consistently outperforms either AI or human judgment alone (39,40). For instance, AI-physician collaborations in prostate cancer diagnostics achieved 90-98% accuracy versus AI alone reaching 86-94% (41–43).

Having outlined the socio-economic impacts of automation, we now turn to the legal frameworks that must govern them. The challenge of responsibility becomes concrete when we consider legal accountability for AI-induced harm. Traditional legal frameworks (both civil and criminal law) assume a clear link between an agent's intent (mens rea) and action (actus reus). AI systems complicate this link because their "actions" may result from complex probabilistic algorithms and data-driven learning processes rather than direct human intent. Consider an autonomous surgical robot that causes an injury due to an unexpected edge case in its programming, or an AI-driven diagnostic system that consistently under-diagnoses a certain demographic due to bias in training data. In such cases, establishing causality and fault is challenging. Multiple factors might be involved: the design of the algorithm, the quality of training data, the way the hospital deployed the system, and even the clinician's oversight (or lack thereof) in using the tool. Legal scholars increasingly suggest that we move toward models of distributed responsibility. Instead of looking for a single culpable party, we might allocate responsibility across the AI's supply chain: the developers, the manufacturers, the regulators who approved it, the users, and the institutions that adopted it. This concept of shared accountability is gaining traction especially for complex AI like self-driving car fleets or clinical AI support systems, where no single actor has full control over outcomes. Regulators worldwide are starting to respond to these challenges. The European Union, for example, has proposed an Artificial Intelligence Act (published in 2024) that uses a risk-based regulatory approach (44). High-risk AI applications will face strict requirements for safety, explainability, human oversight, and fairness. Although lawmakers have so far stopped short of granting legal personhood to AI (a notion widely criticized by ethicists for potentially diluting human accountability), there are active discussions on updating liability laws. Frameworks like the EU's emphasize that AI systems must remain tools and that the humans deploying them retain ultimate responsibility (45). In parallel, standards bodies and professional organizations are issuing guidelines for algorithmic accountability and audit trails, so that when something goes wrong, we can trace the decision process of the AI and identify any points of failure or negligence. Overall, ensuring accountability in AI systems is recognized as crucial not only for justice in individual cases of harm but also for public trust in these technologies.

The simulation of understanding

A key philosophical limit of machine agency is that AI can simulate cognition and even empathy without actually possessing understanding or feelings. John Searle's Chinese Room thought experiment, for example, illustrates that a computer could execute rules to convincingly answer Chinese language questions without understanding a single word of Chinese (46). This implies that no matter how intelligent an AI appears, its "intelligence" may be entirely syntactic, lacking any semantics or consciousness. In real-world robotic systems, we see parallels: a social robot might engage a patient in friendly conversation, yet these systems have no awareness or concern; they are simply following complex programming or learned models. Such simulated empathy, while ethically limited, can nonetheless provide pragmatic benefits. For example, studies have shown that patients may feel more at ease interacting with humanoid robots in hospital settings, especially in roles like companionship, preoperative calming, or physical rehabilitation support, provided a human clinician maintains oversight (27). Used carefully, this simulation of care may enhance the patient experience without misleading them about the machine's true capabilities. Still, it remains a crucial bioethical point that if artificial agents cannot possess ethical subjectivity or true understanding, they cannot bear ethical responsibility for their actions. Responsibility must be traced back through the chain of human designers, developers, and deployers who created and managed the system. This places responsibility squarely on the human designers, developers, and deployers to ensure proper oversight, validation, and accountability (47).

Bioethical implications in healthcare: The medical paradigm

Among the most ethically charged domains affected by AI and robotics is healthcare. Medicine is a paradigm case for bioethics, traditionally governed by principles like autonomy, beneficence, non-maleficence, and justice. The introduction of AI into clinical settings tests these principles in new ways. AI-powered tools are now used for diagnosis (e.g. deep learning algorithms reading radiology scans), prognosis (predictive models for disease outcomes), treatment planning (AI recommending personalized therapies), and even direct intertvention (robot-assisted surgeries or AI-controlled anesthesia) (48). AI-enabled diagnostics can analyze medical images or genomic data faster and in some cases more accurately than human experts. Algorithmic decision-making in medicine appears in forms like triage algorithms prioritizing patients in emergency rooms or risk scores guiding treatment eligibility. Experimental autonomous surgical systems have also been created, suggesting a future where robots might conduct certain procedures with minimal human input (27).

These advances offer substantial benefits such as improved accuracy, consistency, and efficiency in care. However, they also raise serious ethical questions. When an AI system makes a mistake (for example, a misdiagnosis or a treatment recommendation that cause a side effect), assigning liability is difficult. Is the software developer responsible for a flawed algorithm? Is the medical device manufacturer responsible for inadequate validation? Is the healthcare institution at fault for deploying the system, or the clinician for over-relying on it? Such scenarios reveal gaps in our current accountability models (49). They also challenge the principle of informed consent: patients typically consent to treatments administered by human professionals, but do they meaningfully consent to the involvement of AI decision tools, especially if they are not fully aware of them? Patient autonomy could be undermined if, for instance, a hospital mandates the use of an AI diagnostic aid and the patient is not given the option to consent to algorithmic evaluation (27). Therefore, to preserve meaningful consent, each patient should receive and sign an AI-disclosure-and-choice form, which must offer an explicit option to be treated through a human-only workflow.

A concern increasingly raised in the literature is the deskilling of healthcare professionals due to overreliance on AI systems. As automated systems take on more diagnostic and procedural roles, there is a risk that clinicians' expertise may erode over time, potentially compromising care when human intervention becomes necessary (38).

To uphold core bioethical principles, scholars have suggested extending and adapting them for the AI era. For autonomy, this might mean ensuring transparency about when AI is used and giving patients the right to a human second opinion. For beneficence and non-maleficence, it means rigorously testing AI systems for safety and efficacy, analogous to drug trials, and monitoring their performance continually after deployment (since AI can drift or behave unpredictably in changing real-world data environments) (48). Justice in healthcare AI demands scrutinizing these systems for bias: there have been cases where diagnostic algorithms underperform for underrepresented groups (for example, skin cancer detection where AI worked less accurately on darker skin, or symptom-checker apps that cater to male symptom presentations over female) (50). If not addressed, such biases can exacerbate health disparities, violating the ethical commitment to equity.

In medical imaging, in particular, studies have shown how AI systems can inherit and amplify biases present in training data, potentially affecting diagnostic accuracy across demographic groups (51).

Technoethicists like Deborah Johnson and James Moor have advocated for technoethical or Value-Sensitive Design approaches (methods that embed ethical and social values directly into a technology's design, deployment, and use): embedding ethical reflection throughout the entire lifecycle of medical AI, from design and data selection to implementation and user training (52,53). This echoes Hans Jonas's Principle of Responsibility, which urges a precautionary approach to powerful new technologies (54). In practice, a Jonas-inspired view in healthcare AI translates to designing systems with built-in ethical safeguards and fail-safes, and using pilot programs or phased rollouts to catch unforeseen issues before widespread use. For example, an autonomous robot for elder care might include strict constraints to always yield control to human caregivers in ambiguous situations, or an AI surgical tool might be programmed to pause and ask for human intervention if it encounters data outside its trained distribution. Continuous post-deployment auditing is equally crucial: stakeholders must monitor AI in the field and be ready to adjust or withdraw systems that show problematic behavior. Above all, such measures ensure that patient welfare and dignity remain central, guiding ethical decisions about AI's appropriate deployment in healthcare.

Social inequalities, justice and distributive fairness in AI

The benefits of AI and robotics are not experienced uniformly across society. While these technologies promise efficiencies and improvements, they often operate within existing social frameworks that are full of inequality. There is a growing concern that AI systems, if left unchecked, will replicate or even amplify structural injustices related to class, race, gender, geography, and disability. AI reflects the values and biases of its creators and the data it is trained on. In healthcare, for example, if an AI diagnostic tool is developed using data primarily from Western hospitals, it may perform poorly in under-resourced settings or among minority populations, thus widening health disparities (50). Similarly, advanced humanoid robots or assistive AI devices may be available only to wealthy institutions or nations, creating a disparity in who can benefit from technological progress (55). Although AI's advantages remain unevenly distributed, especially in healthcare, targeted deployment could significantly aid developing countries in areas such as agriculture, food safety, and infectious-disease surveillance. To realize these gains, stakeholders must confront the barriers

that slow adoption in low-resource settings, including limited financing, inadequate digital infrastructure, shortages of technical expertise, and the absence of clear regulatory frameworks (56,57).

Building on Pierre Bourdieu's theory of social and cultural capital (58), technological capital refers to the ability to access, control, and effectively use advanced technologies like AI. Those with higher education, technical skills, and institutional support are positioned to gain significantly, whereas marginalized groups may face greater risks, such as job displacement or biased surveillance. Communities with resources leverage AI for improved education and healthcare, while disadvantaged communities might disproportionately endure AI-driven surveillance or poorly tested algorithmic systems. For instance, algorithmic bias in criminal justice risk assessments has led to harsher outcomes for minority defendants in some cases, demonstrating how a tool can entrench societal bias under the guise of objectivity (48).

Structural inequality can also be reinforced through simple technical choices. A rehabilitation exoskeleton that isn't designed for diverse body types effectively excludes certain users; a voice-activated assistant that doesn't recognize non-native accents marginalizes those users; an autonomous vehicle trained on city environments might perform unsafely in rural areas. These reflect whose needs and values were prioritized in development. Addressing this requires ethical commitment to inclusivity in design and deployment. Techniques such as participatory design, where end-users from various backgrounds are involved in the creation of AI systems, can help ensure broader perspectives are considered (59). Likewise, conducting bias audits and fairness tests on AI algorithms (particularly for algorithms deployed in high-stakes domains such as healthcare, credit, or hiring) are essential to detect and mitigate disparate impacts. Developers can integrate this practice by following a standardized audit workflow aligned with, for example, the NIST AI Risk Management Framework (60).

While fairness in AI is often foregrounded in ethical discussions, some scholars argue that diagnostic accuracy should take precedence, especially when improving outcomes within specific populations. They suggest focusing on performance within subgroups rather than overall demographic parity (61). AI in healthcare can pose distinct harms and benefits for disabled individuals. Design flaws, lack of inclusive training data, or assumptions about patient autonomy may result in systems that inadequately serve, or even harm, patients with disabilities (62). Ethical frameworks must explicitly consider disability inclusion in both design and deployment phases.

The spectre of a technologically-driven underclass, sometimes described in Harari's terms as the "useless class," raises bioethical issues around human worth and societal obligation. If a segment of the population is rendered economically "useless" not by lack of merit but by systemic automation, how should society respond? Yuval Noah Harari and others prompt us to consider policies that ensure everyone retains access to the means of a dignified life, such as universal basic income or public-service jobs geared towards human-centric roles that AI cannot fill (for example, jobs emphasizing empathy, creativity, or social interaction) (25). Michel Foucault's notion of biopolitics is also relevant: power in the modern age often works through managing populations; deciding which lives are fostered and which are neglected. In a world mediated by AI, those who control the algorithms have a new form of power to shape life chances. Ethical governance of AI must therefore include equity as a core principle, asking who benefits and who might be harmed by a given application (63).

To move toward a fair distribution of AI's benefits, policy proposals have emerged, including algorithmic impact assessments to evaluate and disclose AI's societal effects (64). Data governance models like data trusts or data dividends can provide communities control over their collective data's value (65). Stronger public investment in AI education and re-skilling programs can help displaced workers find new opportunities. Internationally, UNESCO recommends ethical AI use, emphasizing inclusiveness and peace, advocating shared AI advancements across borders to prevent global gaps (66). Crucially, justice in AI is ongoing, requiring structural awareness in design, deliberately accommodating underrepresented cases, and capping costs of life-saving technologies to ensure affordability (67). Inclusive stakeholder dialogue, from patients to justice groups, is vital to identify subtle ways AI might disadvantage people and devise solutions (59,64).

Toward an information ethics

Given the difficulties of fitting AI into anthropocentric ethical models, some philosophers have proposed new frameworks. Luciano Floridi's theory of Information Ethics is one such approach that evaluates the ethical significance of actions in terms of their impact on the broader information environment, rather than the intentions of a conscious agent (68). In Floridi's view, artificial agents can be considered "ethical agents" in a limited sense based on the consequences of their interactions within a system of information (69). For example, an AI diagnostic tool in healthcare can be judged by how well it upholds values like accuracy, privacy, transparency, and non-maleficence in its operation. The AI doesn't need to understand ethics; rather, we imbue the system with ethical principles and evaluate it by how those principles are realized in practice. This functional approach aligns with emerging trends in machine ethics and ethical-by-design engineering. Developers are increasingly attempting to embed ethical constraints into AI through techniques like value alignment, safety layers, and explainability features. The focus shifts from asking whether a robot has ethical intentions to asking how its behavior can be guided by ethical design and governance. In short, information ethics and related frameworks allow us to treat AI as part of a wider ethical landscape, one that includes humans, machines, and institutions, where normativity is distributed and not solely tied to individual human intent.

Conclusions

Artificial intelligence and robotics have shifted from being peripheral tools to becoming transformative forces that reshape the ethical, ontological, and socio-political dimensions of human life. Their integration into critical areas such as clinical decision-making, transportation, labor, and caregiving requires a reexamination of fundamental bioethical concepts, including personhood, autonomy, ethical agency, and justice. As these technologies increasingly influence decisions with significant human consequences, they reveal a growing gap between our technological capabilities and our ethical and legal preparedness. Issues such as algorithmic opacity, centralized control, and reduced transparency raise urgent concerns about consent, accountability, and the protection of human dignity. In this context, bioethics must move beyond a reactive or specialized role and become a proactive, interdisciplinary framework for addressing both immediate harms and long-term implications. This includes adopting anticipatory ethics, embedding principles like fairness, transparency, accountability, and human-centered design into AI systems, and engaging with diverse ethical traditions. Bioethics should adopt a multigenerational framework that tackles clear, high-probability harms such as algorithmic bias and health-care inequity, while remaining flexible towards lower-probability, long-term AI risks through iterative review and inclusive stakeholder input.

References

1. Mizna S, Arora S, Saluja P, Das G, Alanesi WA. An analytic research and review of the literature on practice of artificial intelligence in healthcare. Eur J Med Res 2025; 30(1): 382.
2. Rehman AU, Li M, Wu B, Ali Y, Rasheed S, Shaheen S, et al. Role of artificial intelligence in revolutionizing drug discovery. Fundam Res 2024 May; 5(3): 1273-1287.
3. Hemmatyar A, Soleymani S, Khosravi-Mashizi M, Saberi A, Shirinzadeh-Dastgiri A, Naseri A, et al. Ethical Considerations in the Use of the da Vinci Surgical System in Modern Surgery. Indian J Surg Oncol 2025.
4. Ienca M, Wangmo T, Jotterand F, Kressig RW, Elger B. Ethical Design of Intelligent Assistive Technologies for Dementia: A Descriptive Review. Sci Eng Ethics. 2018; 24(4): 1035–55.
5. Panahi O. Health in the Age of AI: A Family and Community Focus. Archives of Community and Family Medicine 2025; 8(1):11–20.
6. Santoni de Sio F, van den Hoven J. Meaningful Human Control over Autonomous Systems: A Philosophical Account. Front Robot AI 2018 Feb; 5.
7. Burrell J. How the Machine "Thinks:" Understanding Opacity in Machine Learning Algorithms. SSRN Electronic Journal 2015.
8. Calo R. Robotics and the New Cyberlaw. SSRN Electronic Journal 2014.
9. Law KF, Syropoulos S, Earp BD. Artificial intelligence, existential risk and equity: the need for multigenerational bioethics. J Med Ethics. 2024; 50(12): 799–801.
10. Jecker NS, Atuire CA, Bélisle-Pipon JC, Ravitsky V, Ho A. Stoking fears of AI X-Risk (while forgetting justice here and now). J Med Ethics. 2024; 50(12):827–8.
11. Müller VC. Ethics of Artificial Intelligence and Robotics. In: E. N. Zalta, editor. The Stanford Encyclopedia of Philosophy; 2020.
12. Thomson JJ. The Trolley Problem. The Yale Law Journal Company, Inc. 1985;94(6): 1395–415.
13. Porter Z, Habli I, McDermid J, Kaas M. A principles-based ethics assurance argument pattern for AI and autonomous systems. AI and Ethics 2024; 4(2): 593–616.
14. Manna R, Nath R. Kantian Moral Agency and the Ethics of Artificial Intelligence. Problemos 2021; 100:139–51.
15. Groff R, and Symons J. Is AI Capable of Aristotelian Full Moral Virtue? In: Artificial Dispositions: Investigating Ethical and Metaphysical Issues. Sciendo; 2023. p. 45–58.
16. Onyeukaziri JN. Artificial Intelligence and an Anthropological Ethics of Work: Implications on the Social Teaching of the Church. Religions (Basel) 2024; 15(5): 623.
17. Chen W, Chi W, Ji S, Ye H, Liu J, Jia Y, et al. A survey of autonomous robots and multi-robot navigation: Perception, planning and collaboration. Biomimetic Intelligence and Robotics 2025; 5(2): 100203.
18. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019; 25(1): 44–56.
19. Fuchs C. Dallas Smythe Today - The Audience Commodity, the Digital Labour Debate, Marxist Political Economy and Critical Theory. tripleC: Communication, Capitalism & Critique Open Access Journal for a Global Sustainable Information Society 2012; 10(2): 692–740.
20. Kalberg S. The Protestant Ethic and the Spirit of Capitalism. The Stanford Encyclopedia of Philosophy 2020.
21. Habermas J. The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society. 1st ed. T. McCarthy (Translator), editor. Vol. 1. Boston, MA: Beacon Press; 1984.
22. Rifkin J. The End of Work: The Decline of the Global Labor Force and the Dawn of the Post-Market Era. 1st ed. New York, NY: Putnam Publishing Group; 1995.
23. OECD Economic Outlook, Volume 2023 Issue 1. OECD; 2023.
24. World Economic Forum. The Future of Jobs Report 2020. 2020.
25. Harari YN. Homo Deus: A Brief History of Tomorrow. 1st ed. New York: HarperCollins; 2017.
26. Arendt H. The Human Condition. 1st ed. Chicago, IL: University of Chicago Press; 1958.
27. van Wynsberghe A. Healthcare Robots: Ethics, Design, and Implementation. 1st ed. New York: Routledge; 2016.
28. Maleki Varnosfaderani S, Forouzanfar M. The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century. Bioengineering. 2024; 11(4): 337.
29. Hoppe N, Härting RC, Rahmel A. Potential Benefits of Artificial Intelligence in Healthcare 2023; 225–49.

30. Chustecki M. Benefits and Risks of AI in Health Care: Narrative Review. Interact J Med Res 2024; 13: e53616.
31. Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Reza Afrash M. Generative artificial intelligence in healthcare: A scoping review on benefits, challenges and applications. Int J Med Inform 2024; 188: 105474.
32. Lane WB. Student interactions with generative AI. Nat Phys 2025; 21(6): 866–7.
33. Jin Q, Chen F, Zhou Y, Xu Z, Cheung JM, Chen R, et al. Hidden flaws behind expert-level accuracy of multimodal GPT-4 vision in medicine. NPJ Digit Med 2024; 7(1): 190.
34. Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. J Biomed Inform 2004; 37(3): 193–204.
35. Garrouste-Orgeas M, Philippart F, Bruel C, Max A, Lau N, Misset B. Overview of medical errors and adverse events. Ann Intensive Care 2012; 2(1): 2.
36. Wachter RM, Brynjolfsson E. Will Generative Artificial Intelligence Deliver on Its Promise in Health Care? JAMA 2024; 331(1): 65.
37. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019; 2(1): 77.
38. ElHassan BT, Arabi AA. Ethical forethoughts on the use of artificial intelligence in medicine. International Journal of Ethics and Systems 2025; 41(1): 35–44.
39. Terranova C, Cestonaro C, Fava L, Cinquetti A. AI and professional liability assessment in healthcare. A revolution in legal medicine? Front Med (Lausanne) 2024; 10: 1337335.
40. Baurasien BK, Alareefi HS, Almutairi ¶ Diyanah Bander, Alanazi ¶ Maserah Mubrad, Alhasson ¶ Aseel Hasson, Alshahrani AD, et al. Medical errors and patient safety: Strategies for reducing errors using artificial intelligence. Int J Health Sci (Qassim) 2023; 7(S1): 3471–87.
41. Riaz I Bin, Harmon S, Chen Z, Naqvi SAA, Cheng L. Applications of Artificial Intelligence in Prostate Cancer Care: A Path to Enhanced Efficiency and Outcomes. American Society of Clinical Oncology Educational Book 2024; 44(3): e438516.
42. Padhani AR, Papanikolaou N. AI and human interactions in prostate cancer diagnosis using MRI. Eur Radiol 2025.
43. Arita Y, Roest C, Kwee TC, Paudyal R, Lema-Dopico A, Fransen S, et al. Advancements in artificial intelligence for prostate cancer: Optimizing diagnosis, treatment, and prognostic assessment. Asian J Urol 2025.
44. AI Act. Available at https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai. Accessed on 19/05/2025.
45. European Commission. White Paper on Artificial Intelligence: A European Approach to Excellence and Trust. Brussels; 2020.
46. Searle JR. Minds, brains, and programs. Behavioral and Brain Sciences 1980; 3(3): 417–24.
47. Matarić MJ, Eriksson J, Feil-Seifer DJ, Winstein CJ. Socially assistive robotics for post-stroke rehabilitation. J Neuroeng Rehabil 2007; 4: 5.
48. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019; 6(2): 94–8.
49. Harihara Subramanian G, Verma A. Crowd risk prediction in a spiritually motivated crowd. Saf Sci 2022; 155: 105877.
50. Sasseville M, Ouellet S, Rhéaume C, Couture V, Després P, Paquette JS, et al. Risk of Bias Mitigation for Vulnerable and Diverse Groups in Community-Based Primary Health Care Artificial Intelligence Models: Protocol for a Rapid Review. JMIR Res Protoc. 2023; 12: e46684.
51. Koçak B, Ponsiglione A, Stanzione A, Bluethgen C, Santinha J, Ugga L, et al. Bias in artificial intelligence for medical imaging: fundamentals, detection, avoidance, mitigation, challenges, ethics, and prospects. Diagn Interv Radiol 2025; 31(2): 75-88.
52. Friedman B, Hendry DG. Value Sensitive Design. The MIT Press; 2019.
53. MOOR JH. What is computer ethics?. Metaphilosophy. 1985; 16(4): 266–75.
54. van Wynsberghe A. Service robots, care ethics, and design. Ethics Inf Technol 2016; 18(4): 311–21.
55. Nussbaum MC. Creating Capabilities. Harvard University Press; 2013.
56. Ahmad A, Liew AXW, Venturini F, Kalogeras A, Candiani A, Di Benedetto G, et al. AI can empower agriculture for global food security: challenges and prospects in developing nations. Front Artif Intell 2024; 7.
57. Viswanathan VS, Parmar V, Madabhushi A. Towards equitable AI in oncology. Nat Rev Clin Oncol 2024; 21(8): 628–37.
58. Hannaway J, Richardson JG. Puzzles, Facts, and Tensions: Inquiry in the Sociology of Education. Educational Researcher 1987; 16(6): 43.
59. Robertson T, Simonsen J. Challenges and Opportunities in Contemporary Participatory Design. Design Issues 2012; 28(3): 3–9.
60. AI Risk Management Framework | NIST. Available at https://www.nist.gov/itl/ai-risk-management-framework.
61. Sabuncu MR, Wang AQ, Nguyen M. Ethical Use of Artificial Intelligence in Medical Diagnostics Demands a Focus on Accuracy, Not Fairness. NEJM AI 2025; 2(1).
62. Binkley CE, Reynolds JM, Shuman AG. Health AI poses distinct harms and potential benefits for disabled people. Nat Med 2025; 31(1): 12–3.
63. Amani B. AI and Equality by Design. Artificial Intelligence and the Law in Canada 2021.
64. Whittaker M, Crawford K, Dobbe R, Fried G, Kaziunas E, Mathur V. AI Now Report 2018.
65. Delacroix S, Lawrence ND. Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. Int Data Priv Law 2019.
66. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell 2019; 1(9): 389–99.
67. Himmelreich J, Lim D. Artificial Intelligence and Structural Injustice: Foundations for Equity, Values, and Respon-

sibility 2022.
68. Floridi L. The Ethics of Information . Oxford : Oxford University Press; 2013.
69. Floridi L. Distributed morality in an information society. Sci Eng Ethics. 2013; 19(3): 727–43.